81-2138

26 August 1981

MEMORANDUM FOR: Director of Data Processing
Director of Central Reference

STAT FROM : [redacted]

SUBJECT : Preliminary Study of PRB's Pre-Publication Review Process

REFERENCE : A. Memo to DDA from D/PA (DDA-81-1226), dated 9 June '81, subject PRB Reference Center

B. Memo to D/PA from DDA (ODP-81-7058), same subject

1. Attached are the results of our preliminary investigation of PRB's manuscript collection as agreed in the reference. In summary, we believe there is a requirement for automation and that the collection will lend itself to a computer supported document reference retrieval system.

2. Although our study of the collection and of PRB's retrieval needs was not extensive, we have concluded that a formatted file is the best approach. Using controlled and possibly structured keywords and keyword phrases as the basic retrieval element PRB could maintain good control over their growing inventory. We have not addressed the question of the form of the manuscripts themselves, but only to information extracted from them. The ultimate microforming of the manuscripts could be addressed when, and if, PRB chooses to continue with systems development.

3. Our primary concern with the recommendation of a formatted file is the commitment and resources necessary to support such a system. The strawman we have put together is a first attempt to size the commitment PRB must be willing to make. Both file conversion and current file input and retrieval have been considered. Indeed even with this strawman many other supplementary support items such as verification routines, software maintenance and development, thesaurus construction and indexing procedures and guides have not been addressed. The estimates should be enough, however, to give PRB a glimpse of what would be necessary to implement and maintain the system.

4. In considering the machine support of this application we have discussed and explored various existing software and word processing systems. The existing system that appears to offer the most promise is GIMS although the use of a text processor would be a most unusual and interesting approach.

5. If you agree with our findings and with our recommendation the attached report can be released to PRB.

STAT

26 August 1981

STAT MEMORANDUM FOR: [redacted]
Chief, Public Affairs Branch

FROM : Bruce T. Johnson
Director of Data Processing

SUBJECT : Response to Public Affairs' Request for DDA/ODP Assistance

REFERENCE : A. Memo to DDA from D/PA (DDA-81-1226), dtd. 9 June 1981, subj. PRB Reference Center

B. Memo to D/PA from DDA (ODP-81-7058), Same Subject

1. As agreed to in our 9 June 1981 meeting and documented in the referenced memoranda, a preliminary study of the PRB's information storage and retrieval needs has been completed. The attached paper contains the findings

STAT and recommendations [redacted]

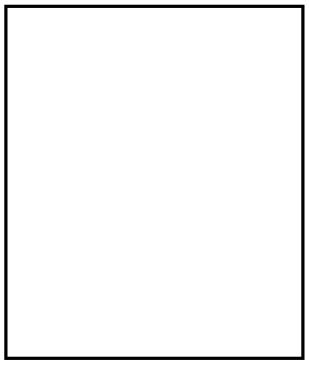
2. Their recommendation of a formatted file approach over a full text retrieval system would significantly reduce the resources required for converting textual manuscripts to machine readable form. Furthermore, if the indexing and abstracting is done well, it will provide the retrieval flexibility needed by PRB. To accomplish this, however, will require a full time data base manager/indexer which will have to come from PRB or Externals Affairs staffing complement. If a comparable position exists or can be created, I recommend we proceed to the requirements gathering step as defined in the attached report.

3. If you have any questions regarding the Preliminary Investigation Report, please call [redacted] STAT

Bruce T. Johnson

Preliminary Investigation Report

for the Publications Review Board



STAT

August 27, 1981

Approved For Release 2003/11/04 : CIA-RDP84-00933R000100120004-6

1. Problem Definition - This preliminary investigation was conducted to determine what approach should be taken in providing an automated system for the storage and retrieval of pertinent information related to Publication Review Board's (PRB) pre-publication review process. The problem as stated by Office of Public Affairs (now Public Affairs Branch) is one of being able to recall what information has been disclosed to the general public through the review mechanism and what information has been withheld.

2. Findings - To begin, we believe that the PRB application is a good candidate for ADP control. The variety and amount of information to be controlled and the need for a timely, systematic organized search and retrieval apparatus supports this belief. Our initial reaction is that it is not a likely candidate for full text processing. Data conversion requirements, the size of the data base to be initially converted (40,000 pages), the projected file growth and storage requirements are the primary reasons for our decision. Eliminating full text processing as an alternative narrows the selection to a formatted file approach, that is, the creation of indexes/records containing information about the manuscripts; the manuscripts themselves being retained in a separate collection.

From a systems point-of-view, the consideration of a formatted file application brings up many points regarding support of the application that should be addressed before a decision to proceed is made. Such an approach will require considerable resources for data reduction, input and file maintenance. It will require a disciplined environment that includes an information abstraction and data entry capability as well as a quality control mechanism. Additionally it could introduce complexities and changes in PRB's office procedures and responsibilities that could affect system design. For example, procedures may have to be established for logging and tracking the manuscript in order to insure that the final disposition has been made and the file record is complete.

In order to assist PRB in analyzing their needs and commitments we have constructed a file resources strawman (attachments 1-5). These estimates are based on a review of a sample of manuscript files currently held in PRB and from initial discussions with PRB personnel.

Using the attached estimates we recommend at least one person fulltime to support current file needs. This estimate presumes this person will have the various skills necessary to perform the functions of control, abstract, input, maintain and retrieve, and a first hand awareness of on-line data entry and ad hoc subject retrieval. Ideally a fully trained and experienced abstractor/indexer would be

desirable. This experience is absolutely necessary to initially maintain the lower range of time estimates and to support a high retrieval/indexing relevancy rate. Of particular concern is the time allocated to data base management functions. At implementation this expenditure will be weighted to the high range figure. Gradually as experience grows and as reference tools are completed the expenditure should ease. About six months will be reqired for this cycle to settle down.

In addition to keeping up with current receipts the conversion of present file holdings is recommended. The conversion of this data base is estimated to require approximately 1/2 manyear. Using the lower resource allocation figure we anticipate this task to complement and to support the current file building operations. It of course will slow down this process unless additional resources are allocated.

The strawman record structure is based on three groups of information about a manuscript - bibliographic data, an abstract of the theme and/or subjects treated and an abstract of the reviewers' comments. Each information group has been described as a subrecord. These subrecords are considered, for the purpose of this file estimate, to be independent for input and maintenance activities. That is, each subrecord may be input to the system as it is completed rather than delaying input until all subrecords are available. Intermittent input allows the system to serve as a control and tracking tool as well as a retrospective retrieval device. Special emphasis on maintenance functions is stressed as each subrecord may be accessed several times to input information as it becomes available; this is especially true in subrecords 1 and 3. At retrieval, however, the record is addressed as a coordinated whole.

3. Recommendations - If based on these data a decision to proceed is made, we would then recommend the formation of a file design team. Composed of a PRB representative, a computer system analyst, and an indexing expert, this team would be responsible for a complete system requirements and file design document. After the requirements have been defined, the group will dissolve and the ODP analyst will write a project proposal for a system to be developed by ODP Applications. This proposal will include all aspects of system design, development, and implementation.

PRB FILE "STRAWMAN"

RECORD STRUCTURE

Each manuscript is represented by one-three part index record. A record is not complete until all three subrecords are input. Each subrecord, however, may be input separately in a unique maintenance action.

Subrecord 1 contains

bibliographic data examples: author's name, title, PRB control number, date submitted, document type, date of comments

comments: basically data data currently controlled in a PRB RAMIS formatted file -- with certain standardizations. (dates, document type, name)

estimated size 400 characters

Subrecord 2 contains

subject abstract (keywords/ keyword phrases) examples: media control, disinformation, CIA field station

STAT

comment: this strawman uses keywords/keyword phrases without additional encoding. The use of codes to represent concepts and/or areas should be considered in future requirements studies. In addition the linkage of areas to keywords/concepts is viewed as a necessary retrieval requirement.

estimated size 750 characters

Subrecord 3 contains

reviewing official's comments and/or concerns (keywords/keyword phrases) examples: operations -

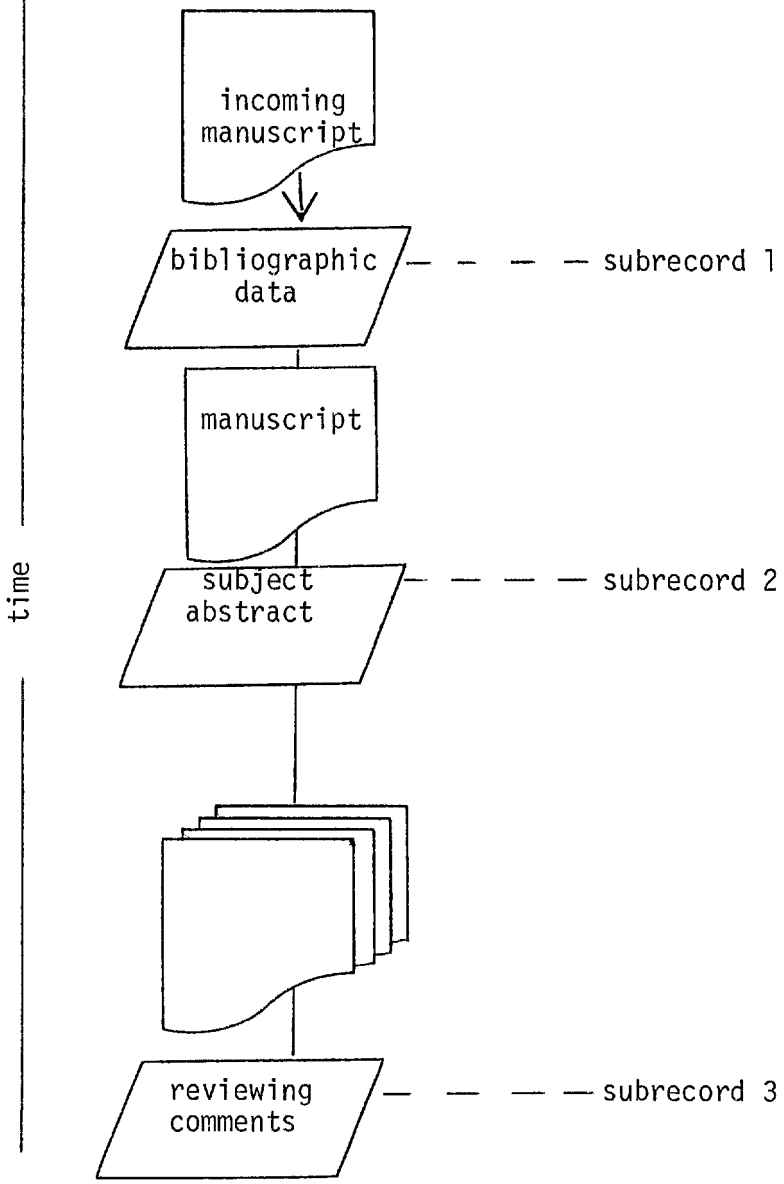
STAT

comment: this strawman uses keywords/keyword phrases without additional encoding. The use of codes to represent concepts and/or areas should be considered in future requirements studies. As in subrecord 2 the linkage of area with the keywords/concepts is most important. The addition of page number to the indexing phrase is an enhancement that may have merit.

estimated size 750 characters

PRB FILE "STRAWMAN"

DOCUMENT FLOW




Approved For Release 2003/11/04 : CIA-RDP84-00933R000100120004-6

PRB FILE STRATEGY

## DATA BASE SIZE AND GROWTH RATE

PRB holdings as of 1 August 1981

| | |
|---|---|
| 109 | books |
| 270 | articles |
| 21 | book reviews |
| 11 | outlines |
| 12 | speeches |
| 27 | other |
| 450 | |

| Distributed Record Size | Greater Manuscripts (Books) | Lesser Manuscripts (Articles, etc.) |
|---|---|---|
| Subrecord 1 | 400 char. | 400 char. |
| Subrecord 2 | 750 char. | 350 char. |
| Subrecord 3 | 750 char. | 350 char. |
| | 1,900 char. | 1,100 char. |

Data Base Size - to be converted (pre CY Aug 81)

| | | |
|---|---|---|
| Books | 109 x 1,900 char. | = 207,100 char. |
| Articles, etc. | 341 x 1,100 char. | = 375,100 char. |
| | TOTAL | = 582,200 char. |

Growth Rate (based on projected CY 81 rate)

| | | |
|---|---|---|
| Books | 24 x 1,900 char. | = 45,600 char. |
| Articles, etc. | 176 x 1,100 char. | = 193,600 char. |
| | TOTAL | = 239,200 char. |

PRB FILE "STRAWMAN"

PRB RESOURCES REQUIRED FOR CURRENT DATA BASE MANAGEMENT
(based on projected CY81 input rate)

| FUNCTION | TYPE OF MANUSCRIPT | TIME REQUIRED/ MANUSCRIPT | X | RATE OF INPUT/YEAR | TOTAL TIME/YEAR |
|---|---|---|---|---|---|
| Bibliographic Indexing | Books | 15-30 min | | 24 | 6 - 12 |
| | Articles | 15-30 min | | 176 | 44 - 88 |
| Abstracting - Subject | Books | 2-4 hrs | | 24 | 48 - 96 |
| | Articles | 30 min - 1 hr | | 176 | 88 - 176 |
| Abstracting - Index reviewers' Comments | Books | 2-4 hrs | | 24 | 48 - 72 |
| | Articles | 15-30 min | | 176 | 44 - 88 |
| Data Entry | Books | 30 min - 1 hr | | 24 | 12 - 24 |
| | Articles | 15-30 min | | 176 | 44 - 88 |
| Data Base Mgt- | --- | 2-3 hrs/day | | --- | 520 - 780 |
| | | | | TOTAL | 854 - 1,424 manhours |

PRB FILE "STRAWMAN"

RESOURCES FOR DATA BASE CONVERSION
(based on current holdings)

| FUNCTION | TYPE OF MANUSCRIPT | TIME REQUIRED/ MANUSCRIPT | X | NUMBER CURRENTLY HELD BY PRB | TOTAL HOURS |
|---|---|---|---|---|---|
| Biographic Indexing | Books | 15 min | | 109 | 27.25 |
| | Articles | 15 min | | 341 | 85.25 |
| Abstracting - Subject | Books | 2 hrs | | 109 | 218 |
| | Articles | 30 min | | 341 | 170.50 |
| Abstracting - Index Reviewers' Comments | Books | 2 hrs | | 109 | 218 |
| | Articles | 15 min | | 341 | 85.25 |
| Data Entry | Books | 30 min | | 109 | 54.50 |
| | Articles | 15 min | | 341 | 85.25 |
| | | | | TOTAL | 944. manhours |